Adaptive Langevin Sampler for Separation of t-distribution Modelled Astrophysical Maps

Koray Kayabol, *Member, IEEE,* Ercan E. Kuruoğlu, *Senior Member, IEEE,* José Luis Sanz, Bülent Sankur, *Senior Member, IEEE,* Emanuele Salerno and Diego Herranz

***Abstract*—We propose to model the image differentials of astrophysical source maps by Student's t-distribution and to use them in the Bayesian source separation method as priors. We introduce an efficient Markov Chain Monte Carlo (MCMC) sampling scheme to unmix the astrophysical sources and describe the derivation details. In this scheme, we use the Langevin stochastic equation for transitions, which enables parallel drawing of random samples from the posterior, and reduces the computation time significantly (by two orders of magnitude). In addition, Student's t-distribution parameters are updated throughout the iterations. The results on astrophysical source separation are assessed with two performance criteria defined in the pixel and the frequency domains.**

***Index Terms*—Bayesian source separation, Multi-channel denoising, Metropolis-Hastings, Langevin stochastic equation, MCMC, Astrophysical images, Student's t-distribution.**

I. Introduction

THE Bayesian framework, which enables the inclusion of prior knowledge in problem formulation, has recently been utilized to improve the performance of Blind Source Separation (BSS) techniques. In the context of image separation, one obvious type of prior information is the spatial [1] or spatio-chromatic [2] dependence among the source pixels.

While there are three conditions that ensure separability of sources, namely, non-Gaussianity, non-whiteness and non-stationarity [3], we choose to exploit spatial correlation (i.e. spatial non-whiteness). The prior densities are constituted by modeling the image differentials in different directions as Multivariate Student's t-distributions [4]. The t-distribution has some convenient properties for our model: If the degree of freedom parameter of the distribution goes to infinity, it approaches a normal density; conversely, if the degree of freedom parameter equals 1, the density becomes Cauchy. Therefore the t-distribution is a flexible and tractable statistical model for data ranging from broad-tailed to normally distributed. The first examples of use of t-distribution in inverse imaging problems can be found in [5] and [6]. In [6], it is shown that the t-distribution approximates the distribution of the wavelet coefficients of an image more accurately. In recent papers, it has been used for image restoration [7] and deconvolution [8]. Notice that the degree of freedom parameter of the t-distribution has the same role as the regularization parameter of the Markov Random Field (MRF) models. The MRF prior with Cauchy density, which was first proposed in [10] for inverse imaging, was used in [1] for source separation. The model used in [1] is an approximation to the t-distribution, which is presented in Section V-A.

The t-distribution has already been used in Bayesian audio source separation [9] to model the discrete cosine transform coefficients of the audio signals. It was reported that the t-distribution prior had improved the sound quality over the finite mixture-of-Gaussians prior. In this study, to solve the Bayesian BSS problem for images without incurring in smoothing artifacts, we propose the t-distribution for modeling the local pixel differences.

We use the joint posterior density of the complete variable set to obtain a joint estimate of all the variables. In this Bayesian approach, the BSS problem can be solved by maximizing the joint posterior density of the sources, the mixing matrix and the source prior model parameters [12], [13]. A method for solving the joint posterior modal estimation problem is the Iterated Conditional Mode (ICM) method, which maximizes the conditional densities sequentially for each variable [11]. If the mode of the conditional density cannot be found analytically, any deterministic optimization method can be used [13]. However, under any non-Gaussian hypothesis, ICM does not guarantee a unique global solution.

Another algorithm suitable for learning the Gaussian MRF is the Expectation-Maximization (EM) method. Using the Mean Field Approximation (MFA), the expectation step of the EM algorithm can be calculated analytically [14]. The MFA under Gaussian model assumption causes smoothing the edges in the image. The reason underlying these smoothing effects is that the Gaussian approximation violates the edge preserving property of the prior density, and the effect is proportional to the amount of noise. The image model with spatially varying variance parameter in variational Bayesian approximation can help overcome the smoothing problem [7], [8]. In [15], [14], deterministic optimization techniques have been used for the MRF. In [1], a Gibbs sampling stochastic optimization procedure is used. Since it is not possible to draw samples in a simple way due to the MRF priors in Gibbs

Manuscript received September 29, 2009; revised March 15, 2010 and accepted March 28, 2010. Koray Kayabol undertook this work with the support of the "ICTP Programme for Training and Research in Italian Laboratories, Trieste, Italy, through a specific operational agreement with CNR-ISTI, Italy. Partial support has also been given by the Italian Space Agency (ASI), under project COFIS (Cosmology and Fundamental Physics). The project is partially supported by CNR-CSIC bilateral project.

K. Kayabol, E. E. Kuruoğlu and E. Salerno are with the ISTI, CNR, via G. Moruzzi 1, 56124, Pisa, Italy, (e-mail: koray.kayabol@isti.cnr.it; ercan.kuruoglu@isti.cnr.it; emanuele.salerno@isti.cnr.it).

J. L. Sanz and D. Herranz are with the IFCA, University of Cantabria, Avda. Los Castros s/n 39005, Santander, Spain, (e-mail: sanz@ifca.unican.es; herranz@ifca.unican.es).

B. Sankur is with the Bogazici University, Electrical & Electronics Eng. Dept., 34342, Bebek, Istanbul, Turkey, (e-mail: bulent.sankur@boun.edu.tr).

distribution form, a Metropolis embedded Gibbs sampling has been adopted. In [1], it is reported that the Monte Carlo results with less image smoothing, but the cost of avoiding smoothing artifacts is a significant increase in the convergence time.

In this study, we propose a more efficient Monte Carlo Markov Chain (MCMC) sampling method in lieu of random walk Metropolis. To produce proposal samples in parallel, we resort to the Langevin stochastic equation [16], [17], [18], while the proposed samples are accepted or rejected by the Metropolis scheme. In statistical physics, the Langevin equation [16] is used to describe the Brownian motion of particles in a potential field and has been used to obtain a smart MC algorithm in [17]. Another parallel sampling algorithm is the Hamilton Monte Carlo, which is the generalized version of the Langevin sampler [18]. We conjecture that, with the samples produced in parallel by the Langevin equation, the convergence time of the algorithm will be significantly reduced.

Parameter estimation in Bayesian edge preserving inverse imaging problems with Gibbs distributions is a troublesome process because of the partition function. Although there are some methods [19], [20] that calculate the parameters using Monte Carlo techniques, their computational burdens are prohibitive. One can resort to the Pseudo Likelihood (PL) approximation to make the partition function separable, which is more convenient for parameter estimation with the Maximum Likelihood (ML) method. In [21], two Bayesian approaches have been used to estimate parameters, namely, the Maximum-a-Posteriori (MAP) and evidence approaches. An approach to estimate the regularization parameter from the PL approximation has been recently proposed [22]. The multivariate t-distribution is also a PL approximation to MRF and has advantages over MRF in parameter calculations.

There are two types of parameters in edge preserving inverse imaging. The first one is the adaptive edge preserving parameter, which is also known as threshold parameter. We interpret the threshold parameter as the scale parameter of the t-distribution. The other parameter is the regularization parameter, which adjusts the balance between the likelihood and the prior. The regularization parameter corresponds to the degree of freedom (dof) parameter of the t-distribution. To estimate the scale and dof parameters of our t-distribution, we use ML estimation via EM algorithm as in [30]. A similar approach has been used in [7] and [8].

In a comparative study among image source separation algorithms [1], we have found that the Bayesian formulation with MRF prior and Gibbs sampling outperformed the heuristic and the other Bayesian approaches. This work is based on [1], and aims to achieve a much faster MCMC implementation without compromising its good performance. With this goal in mind, we have been testing our algorithms on a current problem of modern astrophysics: the separation of radiation source maps from multichannel images of the sky at microwave frequencies. In particular, we have been applying our algorithm to the separation of the Cosmic Microwave Background (CMB) radiation from the galactic emission (synchrotron and thermal dust emissions) using realistically simulated sky maps.

The original contributions of the paper hinge on two aspects. First, we adopt the t-distribution to build a prior model of the source maps. More specifically, the t-distribution is used to model the differentials of the sources, as done by the first-order homogeneous MRF models. This is advantageous because the flexibility of the t-distribution allows each source differential to assume a different model whether impulsive or Gaussian, simply by setting the dof parameter. The second contribution is the introduction of the Langevin sampling scheme in lieu of the random walk Gibbs sampling. As opposed to pixel-by-pixel sampling, Langevin sampling generates the samples in parallel, thus leading to much faster convergence. Furthermore, the samples are drawn in an informed way, since their generation follows the gradient descent direction on an energy surface.

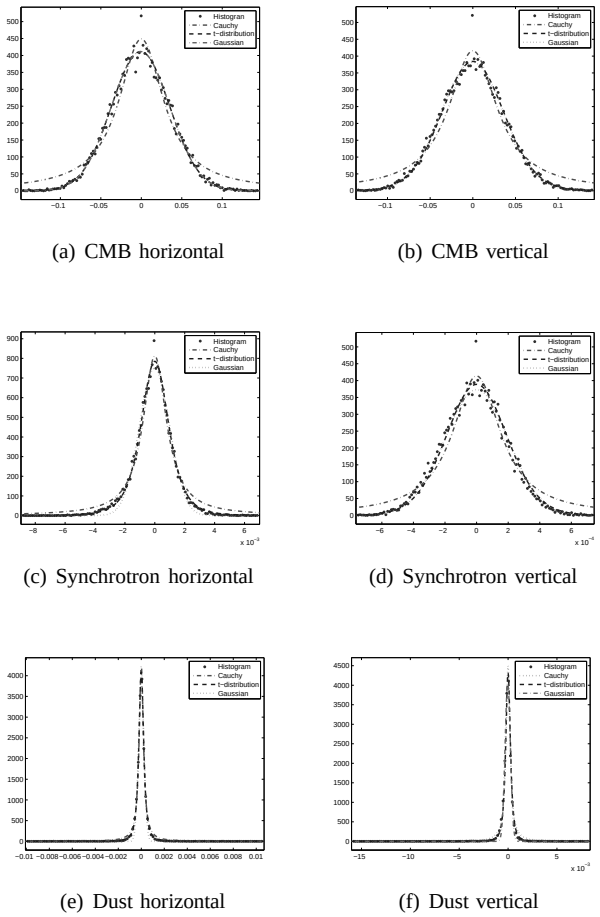
Section II is a brief introduction to astrophysical sources. In Section III, the component separation problem in observational astrophysics is stated. Section IV lays out the Bayesian formulation of the problem. Section V presents the derivation steps of the adaptive Langevin sampler algorithm along with the EM parameter estimation method. The simulation results are presented in Section VI and interpreted in Section VII.

II. An Introduction to Astrophysical Sources

Here, we only give a brief description of the astrophysical radiations considered, referring the interested reader to [23] for details. We are interested in the frequency range 30 to 1000 GHz where the dominant diffuse radiations are the CMB, the galactic synchrotron radiation and the thermal emission from galactic dust. Studying this radiation would help us to understand the distribution and the features of interstellar dust in our galaxy.

The most interesting astrophysical source in the microwave region of the electromagnetic spectrum is the CMB, a relic radiation originating from the time when the Universe was 300.000 years old. The discovery of the CMB is one of the fundamental milestones of modern cosmology and its study allows us to determine fundamental parameters such as the age of the universe, its matter and energy composition, its geometry and many other relevant cosmological parameters. CMB should be a blackbody radiation at a temperature of 2.726 K, thus its emission spectrum should be perfectly known. The CMB emission dominates over the other sources at frequencies around 100 GHz. The CMB temperature is not perfectly anisotropic. Standard cosmological models predict that the CMB anisotropy is Gaussian distributed, although some alternative models permit a certain degree of non-Gaussianity. Current observations are compatible with the hypothesis of Gaussianity. Two all-sky surveys have been made on CMB so far, by NASA's satellites COBE [32] and WMAP [33]. A European mission whose data will be highly accurate and spatially resolved, Planck [23], is about to provide its first full-sky coverage maps.

The CMB signal is mixed with other astrophysical sources of electromagnetic radiation. Relativistic electrons being accelerated by magnetic fields in the Galaxy give rise to synchrotron emission, which dominates over the CMB in regions close to the Galactic plane especially at low frequencies (< 200



(a) CMB horizontal (b) CMB vertical

(c) Synchrotron horizontal (d) Synchrotron vertical

(e) Dust horizontal (f) Dust vertical

Fig. 1. Fitting plots of the image differential histograms (dots) of CMB, synchrotron and dust components in horizontal and vertical directions. The fitted functions proportional to Gaussian (dot line), Cauchy (dash-dot line) and t-distribution (dash line).

GHz). According to observations in other frequency bands, the synchrotron emission spectrum follows a power law with a negative exponent, whose value is presently known with high uncertainty. Synchrotron is the dominant radiation in the low-frequency bands of our range of interest. Inter-stellar dust grains are heated by nearby stars and re-emit thermal radiation in the far infrared region of the electromagnetic spectrum. Dust radiation is dominant in the high end of our range. In particular, it is almost the only significant contribution to the total diffuse radiation between 800 GHz and 1000 GHz. Its emission spectrum should follow a greybody law, with unknown spectral index and an additional degree of freedom given by the thermodynamical temperature of the dust grains. For a short review on CMB astronomy, see [24].

There are other astrophysical sources present at microwave frequencies, such as free-free emission due to free electrons, anomalous dust emission and radiation coming from extra-galactic sources, but their relevance is smaller. In this work we will focus on the main three astrophysical sources present in CMB experiments: CMB, synchrotron and dust.

In order to justify our adoption of the Student's t-distribution in a relevant case, we have selected a $15^\circ \times 15^\circ$ sky patch, located at 0° galactic longitude and 40° galactic latitude, discretized into a 512×512-pixel map. Within this patch, we have introduced simulated CMB, synchrotron, and dust radiation maps (as in [26]). We have computed the source image differentials for horizontal and vertical directions, and estimated their empirical distributions. We have fitted three different functions to the empirical distributions of the image differentials of the astrophysical sources with nonlinear least square method using the Curve Fitting Toolbox of MATLAB. These functions are proportional to Gaussian, Cauchy and t-distribution. Fig. 1 shows the fitting results for CMB, synchrotron and dust images. Table I lists the residual Root Mean Square Errors (RMSE) of the fits. The Gaussian gives the best fit for CMB because the CMB is theoretically distributed as a Gaussian [24]. Overall, the t-distribution appears to be a good choice for modeling the image differential statistics in horizontal and vertical directions of all the components. The estimated dof parameters of t-distributions show that indeed the proposed model assumes from impulsive to Gaussian characteristic underlying each component. If the component is Gaussian as CMB, the dof parameter becomes bigger and if it is impulsive, the dof parameter becomes very small.

TABLE I
ROOT MEAN SQUARE ERROR (RMSE) OF FITTING OF THE IMAGE DIFFERENTIAL HISTOGRAMS OF CMB, SYNCHROTRON AND DUST COMPONENTS. THE LAST COLUMN SHOWS THE ESTIMATED DOF PARAMETERS OF THE t-DISTRIBUTION.

Horizontal direction

	Gaussian	Cauchy	t-distribution	dof
CMB	15.11	30.84	15.47	25.71
Synchrotron	23.70	30.06	15.70	3.59
Dust	67.99	33.21	13.84	1.81

Vertical direction

	Gaussian	Cauchy	t-distribution	dof
CMB	15.75	29.70	15.83	15.75
Synchrotron	19.11	31.00	18.39	19.11
Dust	63.73	68.19	54.42	2.18

III. COMPONENT SEPARATION PROBLEM IN OBSERVATIONAL ASTROPHYSICS

Virtually any application in observational astrophysics has to do with problems of component separation. Indeed, all the astrophysical observations result from the superposition of the radiation sources placed along the line of sight. While very distant sources can be distinguished by the redshift analysis, for nearby sources this is not possible. In fact, physically distinct sources can sometimes be found within a close range of each other. Furthermore, high sensitivity and high resolution measurements can give rise to source mixing problems even in the cases where the radiation under study is dominant over interfering radiations. Apart from redshift analysis, useful methods to distinguish between superimposed physically different radiations include spectral analysis and morphological analysis. In this paper, we only treat the former approach, exploiting the differences in the emission spectra shown by physically distinct radiation sources. This implies that the separation must be done on the basis of measurements made at different frequency bands.

We assume that the observed images, $y_k, k \in \{1, 2, \ldots, K\}$, are linear combinations of L source images. Let the kth observed image be denoted by $y_{k,i}$, where $i \in \{1, 2, \ldots, N\}$ represents the lexicographically ordered pixel index. The image separation problem consists in finding L independent sources from K different observations. If $\mathbf{s}_l$ and $\mathbf{y}_k$ denote $N \times 1$ vector representations of source and observation images, respectively, then the observation model can be written as

$$\mathbf{y}_k = \sum_{l=1}^{L} a_{k,l}\mathbf{s}_l + \mathbf{n}_k, \qquad k = 1, \ldots, K \tag{1}$$

where $\mathbf{n}_k$ is an iid zero-mean noise vector with $\Sigma = \sigma_k^2 \mathbf{I}_N$ covariance matrix and $\mathbf{I}_N$ is an identity matrix. Although the noise is not necessarily homogeneous in the astrophysical maps, in this study we assume that the noise variance is homogeneous within each sky patch and is also known.

Since the observation noise is assumed to be independent and identically distributed zero-mean Gaussian at each pixel, the likelihood is expressed as

$$p(\mathbf{y}_{1:K}|\mathbf{s}_{1:L}, \mathbf{A}) \propto \prod_{k=1}^{K} \exp\left\{-W(\mathbf{s}_{1:L}|\mathbf{y}_k, \mathbf{A}, \sigma_k^2)\right\} \tag{2}$$

$$W(\mathbf{s}_{1:L}|\mathbf{y}_k, \mathbf{A}, \sigma_k^2) = \frac{||(\mathbf{y}_k - \sum_{l=1}^{L} a_{k,l}\mathbf{s}_l)||^2}{2\sigma_k^2} \tag{3}$$

where the mixing matrix $\mathbf{A}$ contains all the mixing coefficients $a_{k,l}$ introduced in (1).

For many purposes, a mixing model of the type (1) is considered to fit reasonably well to an astrophysical observation. The details on how to get an equation similar to (1) from the physics of the problem can be found in [25]. Here, we only summarize the main assumptions made with this purpose. First, we assume that the superposition of the signals originating from different sources is linear and instantaneous. In the astrophysical case, this assumption is clear, since the physical quantities to be measured are superpositions of electromagnetic waves coming, for any bearing, exactly from the same line of sight without any scattering or diffraction effect. The second assumption in modeling astrophysical observations is that each source has an emission spectrum that does not vary with the bearing. This assumption implies that individual radiations result from the product of a fixed spatial template and an isotropic emission spectrum. Both assumptions need closer attention. The precise emission spectrum generated by any physical phenomenon depends on many quantities that may not all be distributed uniformly in the sky. Although in many applications the isotropy assumption has been adopted successfully, in many other cases the space-variability of the radiation sources must be taken into account to allow a good separation to be performed. Furthermore, if the effect of the telescope is taken into account, then the instantaneous model is no more valid since, for the finite aperture, the light captured in a fixed direction in the telescope does not come from that direction alone. In formulas, model (1) becomes

$$\tilde{\mathbf{y}}_k = \mathbf{h}_k * \mathbf{y}_k = \mathbf{h}_k * \sum_{l=1}^{L} a_{k,l}\mathbf{s}_l + \mathbf{n}_k \tag{4}$$

where the asterisk means convolution, and $\mathbf{h}_k$ is the telescope radiation pattern in the k'th observation channel. Note that, if $\mathbf{h}_k$ is the same for all the channels, (4) can be written as

$$\tilde{\mathbf{y}}_k = \sum_{l=1}^{L} a_{k,l}\mathbf{h} * \mathbf{s}_l + \mathbf{n}_k = \sum_{l=1}^{L} a_{k,l}\tilde{\mathbf{s}}_l + \mathbf{n}_k \tag{5}$$

and the problem is again instantaneous for the modified sources $\tilde{\mathbf{s}}_l$, which are the physical sources smoothed by the common radiation pattern $\mathbf{h}$. Unfortunately, especially in the radio- to millimeter-wave ranges, the telescope aperture depends strongly on frequency, and model (5) cannot be adopted directly, unless the observed signals are preprocessed to reduce their angular resolution to the worst available (see [26]).

Since no imaging system can achieve an infinite resolution, we should use (4) as our generative model. However, to avoid problems with the convolutive mixtures, we assume to have a telescope with the same radiation pattern in all the channels, so as to be able to use model (5). Note that this can always be obtained by preprocessing, provided that all the beam patterns are known. Hereafter, as it will not cause any ambiguity, we drop the tilde accent from the symbols used to denote the data and the source vectors.

IV. Source Separation Defined in the Bayesian Framework

A. Source Model

Neighbor pixels in our images have strong dependency. This is demonstrated, for example, in Fig. 2(d) where the scatter-plot shows the first order right neighbor pixels of the dust map shown in Fig. 2(a). The dependency decreases in the high intensity region. This region in the scatter-plot corresponds to spatially localized structures with high image intensity of the map in Fig. 2(a). The existence of a small number of point-like structures in the maps indicates that the dependency assumption is valid. In view of this, we can write an auto-regressive source model using the first order neighbors of the pixel:

$$\mathbf{s}_l = \alpha_{l,d}\mathbf{G}_d\mathbf{s}_l + \mathbf{t}_{l,d} \tag{6}$$

where $d \in \{1, \ldots, D\}$ denotes one of the main directions (left, right, up and down) and $D = 4$ is the cardinality of the set of image differential directions. Matrix $\mathbf{G}_d$ is a linear one-pixel shift operator in direction d, $\alpha_{l,d}$ is the regression coefficient and the regression error $\mathbf{t}_{l,d}$ is an iid t-distributed zero-mean vector with dof parameter $\beta_{l,d}$ and scale parameters $\delta_{l,d}$, $\mathcal{T}(\mathbf{t}_{l,d}|0, \delta_{l,d}\mathbf{I}_N, \beta_{l,d})$. We can justify the iid assumption of $\mathbf{t}_{l,d}$ by plotting (see Fig. 2(e)) the values in $\mathbf{t}_{l,d}$ versus its first order neighbors, $\mathbf{G}_d\mathbf{t}_{l,d}$. We can interpret $\mathbf{t}_{l,d}$ as a decorrelated version of $\mathbf{s}_l$. Fig. 2(b) shows $\mathbf{t}_{l,d}$ for $d = 1$. By comparing Fig. 2(d) and (e), we can say that $\mathbf{t}_{l,d}$ is spatially more independent than $\mathbf{s}_l$.

If the image $\mathbf{s}_l$ were Gaussian distributed, then the regression error would also be Gaussian. However in real images, the regression error is better modelled by some heavy-tailed distribution. The t-distribution can conveniently model the statistics of data whose distribution ranges from Cauchy to Gaussian, and therefore it is a convenient model for the

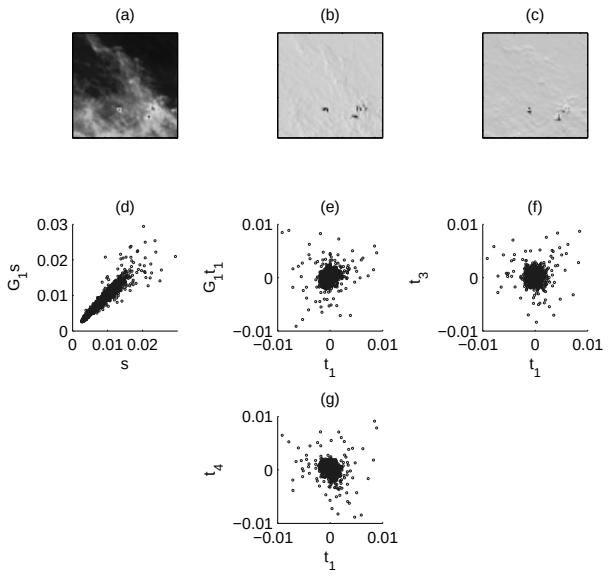


Fig. 2. (a): The synchrotron map $\mathbf{s}$, (b): Right $\mathbf{t}_1 = \mathbf{s} - \mathbf{G}_1\mathbf{s}$ and (c): Up $\mathbf{t}_3 = \mathbf{s} - \mathbf{G}_3\mathbf{s}$ difference maps of (a). (d): Scatter-plot of the first order right neighbor pixels of (a), $\mathbf{s}$ and $\mathbf{G}_1\mathbf{s}$. (e): Scatter-plot of the first order right neighbor pixels of (b), $\mathbf{t}_1$ and $\mathbf{G}_1\mathbf{t}_1$. (f): Scatter-plot of the right difference (b) versus up difference (c), $\mathbf{t}_1$ and $\mathbf{t}_3$. (g): Scatter-plot of the right difference (b) versus left difference (c), $\mathbf{t}_1$ and $\mathbf{t}_4$.

statistics of the high spatial frequency contents of images, such as regression errors [6]. The scale parameter $\delta_{l,d}$ can be assumed as a space-varying parameter to model the highly non-stationary sources, i.e. sparse sources, but homogenous variance assumption has been observed to be adequate for diffuse astrophysical source images for the image patch sizes that we use in the simulations. We use a homogeneous variance since, in our case, the increased complexity derived from inhomogeneity is not justified by a significant improvement in performance.

The regression error $\mathbf{t}_{l,d}$ represents the directional image differential in the direction d. The multivariate probability density function of an image modelled by a t-distribution can be defined as

$$p(\mathbf{t}_{l,d}|\alpha_{l,d},\beta_{l,d},\delta_{l,d}) = \frac{\Gamma((N+\beta_{l,d})/2)}{\Gamma(\beta_{l,d}/2)(\pi\beta_{l,d}\delta_{l,d})^{N/2}} \times \left[1+\frac{\phi_d(\mathbf{s}_l,\alpha_{l,d})}{\beta_{l,d}\delta_{l,d}}\right]^{-(N+\beta_{l,d})/2} \quad (7)$$

where $\phi_d(\mathbf{s}_l,\alpha_{l,d}) = ||\mathbf{t}_{l,d}||^2 = ||\mathbf{s}_l - \alpha_{l,d}\mathbf{G}_d\mathbf{s}_l||^2$ and $\Gamma(.)$ is the Gamma function. We can write the density of $\mathbf{s}_l$ by using the image differentials in different directions, assuming directional independence, as $p(\mathbf{s}_l|\Theta) = \prod_{d=1}^{D} p(\mathbf{t}_{l,d}|\alpha_{l,d},\beta_{l,d},\delta_{l,d})$ where $\Theta = \{\alpha_{1:L,1:D},\beta_{1:L,1:D},\delta_{1:L,1:D}\}$. We can simply justify this assumption by plotting the horizontal regression error versus the vertical one as shown in Fig. 2(f). We can observe in this figure that the scatter plot of the regression errors in left and up directions is almost circular and this justifies our assumption of independence. Fig. 2(g) shows the scatter-plot of right difference versus left difference. In spite of a weak dependence between right and left difference images, we maintain the independency assumption to constitute the products of the regression error probabilities model. A similar approach to constitute a single prior by multiplying different individual priors can be found in [27]. In this study, we use the left/right and up/down differences jointly in the prior model to balance their contributions. Otherwise, we could obtain directionally biased results.

The t-distribution can be written in implicit form by using a Gaussian and a Gamma density. The t-distribution has a dof parameter β, which is itself governed by Gamma distribution with parameter $\beta/2$. The t-distribution has the following form [30]:

$$p(\mathbf{t}_{l,d}|\alpha_{l,d},\beta_{l,d},\delta_{l,d}) = \int p(\mathbf{t}_{l,d}|\nu_{l,d},\delta_{l,d})p(\nu_{l,d}|\beta_{l,d})d\nu_{l,d} = \int \mathcal{N}\left(\mathbf{t}_{l,d}|0,\frac{\delta_{l,d}\mathbf{I}_N}{\nu_{l,d}}\right)\mathcal{G}\left(\nu_{l,d}|\frac{\beta_{l,d}}{2},\frac{\beta_{l,d}}{2}\right)d\nu_{l,d}. \quad (8)$$

The interpretation of this equation is that if $\mathbf{t}_{l,d}|\delta_{l,d},\nu_{l,d}$ is distributed with a normal density $\mathcal{N}(\mathbf{t}_{l,d}|0,\delta_{l,d}\mathbf{I}_N/\nu_{l,d})$ and the parameter $\nu_{l,d}$ has a Gamma prior as $\mathcal{G}(\nu_{l,d}|\beta_{l,d}/2,\beta_{l,d}/2)$, the distribution of $\mathbf{t}_{l,d}|\beta_{l,d},\delta_{l,d}$ becomes a t-distribution such that $\mathcal{T}(\mathbf{t}_{l,d}|0,\delta_{l,d}\mathbf{I}_N,\beta_{l,d})$. The representation in (8) is a particular case of the Gaussian Scale Mixture (GSM) densities. The ML estimation of the parameters $\alpha_{l,d}$, $\beta_{l,d}$ and $\delta_{l,d}$ using the EM method [30] is given in Section V-B.

B. Posteriors

The joint posterior density of all the unknowns in the BSS problem can be written as:

$$p(\mathbf{s}_{1:L},\mathbf{A},\Theta|\mathbf{y}_{1:K}) \propto p(\mathbf{y}_{1:K}|\mathbf{s}_{1:L},\mathbf{A})p(\mathbf{s}_{1:L},\mathbf{A},\Theta) \quad (9)$$

where $p(\mathbf{y}_{1:K}|\mathbf{s}_{1:L},\mathbf{A})$ is the likelihood and $p(\mathbf{s}_{1:L},\mathbf{A},\Theta)$ is the joint prior density of unknowns. The joint prior can be factorized as $p(\mathbf{s}_{1:L}|\alpha_{1:L,1:D},\beta_{1:L,1:D},\delta_{1:L,1:D})\ p(\mathbf{A})\ p(\beta_{1:L,1:D})\ p(\delta_{1:L,1:D})\ p(\alpha_{1:L,1:D})$. Furthermore, since the sources are assumed to be independent, the joint probability density of the sources is also factorized as $p(\mathbf{s}_{1:L}|\Theta) = \prod_{l=1}^{L} p(\mathbf{s}_l|\Theta)$.

Mathematically, we can assume uniform priors for $\alpha_{l,d} \in (-1,1)$, $\delta_{l,d} \in (0,\infty)$ and $a_{k,l} \in (0,\infty)$, because $a_{k,l}$'s are always positive. The practical usage of these priors is explained in Section V-D. We use a conjugate Gamma prior for $\beta_{l,d} \sim \mathcal{G}(1/2, 2\times10^{-3})$. We have determined the respective parameters experimentally. The conditional posteriors of all model parameters are written as

$$\begin{aligned}
p(a_{k,l}|\mathbf{y}_{1:K},\mathbf{s}_{1:L},\mathbf{A}_{-a_{k,l}},\Theta) &\propto p(\mathbf{y}_{1:K}|\mathbf{s}_{1:L},\mathbf{A})\\
p(\alpha_{l,d}|\mathbf{y}_{1:K},\mathbf{s}_{1:L},\mathbf{A},\Theta_{-\alpha_{l,d}}) &\propto p(\mathbf{t}_{l,d}|\Theta)\\
p(\beta_{l,d}|\mathbf{y}_{1:K},\mathbf{s}_{1:L},\mathbf{A},\Theta_{-\beta_{l,d}}) &\propto p(\mathbf{t}_{l,d}|\Theta)p(\beta_{l,d})\\
p(\delta_{l,d}|\mathbf{y}_{1:K},\mathbf{s}_{1:L},\mathbf{A},\Theta_{-\delta_{l,d}}) &\propto p(\mathbf{t}_{l,d}|\Theta)p(\delta_{l,d})\\
p(\mathbf{s}_l|\mathbf{y}_{1:K},\mathbf{s}_{(1:L)-l},\mathbf{A},\Theta) &\propto p(\mathbf{y}_{1:K}|\mathbf{s}_{1:L},\mathbf{A})p(\mathbf{s}_l|\Theta)
\end{aligned} \quad (10)$$

where $-variable$ expressions in the subscripts denote the removal of that variable from the variable set. The parameters α, β and δ have size $L\times D$, $\mathbf{A}$ has size $K\times L$ and the sources have size $L\times N$. Overall there are $(3D+K+N)L$ unknowns.

For parameters $\alpha_{l,d}$, $\delta_{l,d}$ and $\beta_{l,d}$, we exploit the EM method. To estimate the source images, we use a version of the posterior $p(\mathbf{s}_l|.)$ augmented by auxiliary variables and find the estimation with a Langevin sampler. The details are given in Section V.

V. Estimation of Sources and Parameters

In this section, we give the details of the estimation of the sources and the parameters.

A. Sources

We modify the posterior densities of the source images $p(\mathbf{s}_l|\Theta)$ to obtain a more efficient MCMC sampler. In the classical MCMC schemes, a random walk process is used to produce the proposal samples. Although random walk is simple, it affects adversely the convergence time. The random walk process only uses the previous sample for producing a new proposal. Instead of a random walk, we use the Langevin stochastic equation, which exploits the gradient information of the energy function to produce a new proposal. Since the gradient directs the proposed samples towards the mode, the final sample set comes mostly from around the mode of the posterior [28], [29].

The Langevin equation can be obtained from the total energy function. We first define everything in continuous time to give the derivation steps of the Langevin equation, then we transfer them into discrete time. To obtain the total energy function, we introduce a velocity parameter $\mathbf{v}_l(t) = d\mathbf{s}_l(t)/dt$ to define the kinetic energy such that

$$K(\mathbf{v}_l(t)|\mathbf{M}_l) = \frac{1}{2}\mathbf{v}_l^T(t)\mathbf{M}_l\mathbf{v}_l(t) \tag{11}$$

where $\mathbf{M}$ is a diagonal matrix whose diagonal elements correspond to mass parameters $m_{l,n}$ for pixel index $n = 1,\ldots,N$. Using the velocity parameter $\mathbf{v}_l$, the modified version of the posterior density in (10) is written as $p(\mathbf{s}_l,\mathbf{v}_l|\mathbf{y}_{1:K},\mathbf{s}_{(1:L)-l},\mathbf{A},\Theta,\mathbf{M}_l) \propto p(\mathbf{y}_{1:K}|\mathbf{s}_{1:L},\mathbf{A})\, p(\mathbf{s}_l|\Theta)p(\mathbf{v}_l|\mathbf{M}_l)$. More explicitly, it can be written as

$$\begin{aligned} &p(\mathbf{s}_l,\mathbf{v}_l|\mathbf{y}_{1:K},\mathbf{s}_{(1:L)-l},\mathbf{A},\Theta,\mathbf{M}) \propto \\ &\exp\{-(W(\mathbf{s}_{1:L}|\mathbf{A}) + U(\mathbf{s}_l|\Theta) + K(\mathbf{v}_l|\mathbf{M}))\} \end{aligned} \tag{12}$$

where the energy function $U(\mathbf{s}_l|\Theta)$ of a source image can be written in terms of image differentials $\mathbf{t}_{l,d}$ as

$$U(\mathbf{s}_l|\Theta) = \sum_{d=1}^{D} \rho(\mathbf{t}_{l,d}|\Theta). \tag{13}$$

where the function $\rho(\mathbf{t}_{l,d}|\Theta)$ is proportional to the negative logarithm of the t-distribution in (7), that is,

$$\rho(\mathbf{t}_{l,d}|\Theta) = \frac{N+\beta_{l,d}}{2}\log\left[1 + \frac{\phi_d(\mathbf{s}_l,\alpha_{l,d})}{\beta_{l,d}\delta_{l,d}}\right] \tag{14}$$

and the function $\log[1+\phi_d(\mathbf{s}_l,\alpha_{l,d})/\beta_{l,d}\delta_{l,d}]$ is the regularization function proposed in [10]. The terms $(N+\beta_{l,d})/2$ and $\beta_{l,d}\delta_{l,d}$ correspond to the regularization and the threshold parameters, respectively, used in edge preserving image reconstruction.

The energy function $W(\mathbf{s}_{1:L}|\mathbf{A})$ was defined in (3). The total energy function is proportional to the negative logarithm of the posterior. In summary, the three terms correspond, respectively, to the fit to data and to the inertial and the kinetic energy terms. We can define the Lagrangian function: $L(\mathbf{s}_l(t),\mathbf{v}_l(t)) = K(\mathbf{v}_l) - W(\mathbf{s}_{1:L}) - U(\mathbf{s}_l)$ and write the Lagrange-Euler equation for the Lagrangian as follows

$$\frac{d}{dt}\left(\frac{\partial L(\mathbf{s}_l(t),\mathbf{v}_l(t))}{\partial \mathbf{v}_l}\right) = \frac{\partial L(\mathbf{s}_l(t),\mathbf{v}_l(t))}{\partial \mathbf{s}_l}, \tag{15}$$

$$\mathbf{M}_l\frac{d\mathbf{v}_l}{dt} = -\frac{\partial}{\partial \mathbf{s}_l}E(\mathbf{s}_l). \tag{16}$$

where $E(\mathbf{s}_{1:L}) = W(\mathbf{s}_{1:L}) + U(\mathbf{s}_l)$. If we discretize the dynamics in (16) and velocity $\mathbf{v}_l(t)$ using the Leapfrog method [18], we obtain the following three-step iteration

$$\mathbf{v}_l^{k+\frac{1}{2}} = \mathbf{v}_l^k - \frac{1}{2}\tau_l\mathbf{M}_l^{-\frac{1}{2}}\mathbf{g}(\mathbf{s}_{1:L}^k) \tag{17}$$

$$\mathbf{s}_l^{k+1} = \mathbf{s}_l^k + \tau_l\mathbf{v}_l^{k+\frac{1}{2}} \tag{18}$$

$$\mathbf{v}_l^{k+1} = \mathbf{v}_l^{k+\frac{1}{2}} - \frac{1}{2}\tau_l\mathbf{M}_l^{-\frac{1}{2}}\mathbf{g}(\mathbf{s}_{1:L}^k) \tag{19}$$

where $\mathbf{g}(\mathbf{s}_{1:L}^k) = [\nabla_{\mathbf{s}_l}E(\mathbf{s}_{1:L})]_{\mathbf{s}_{1:L}=\mathbf{s}_{1:L}^k}$, $\nabla_{\mathbf{s}_l}$ is the gradient with respect to $\mathbf{s}_l$ and τ_l is the discrete time step. If we define a diagonal matrix $\mathbf{D}_l^{\frac{1}{2}} = \tau_l\mathbf{M}_l^{-\frac{1}{2}}$, so that, for the nth pixel, the diffusion coefficient is $\mathbf{D}_l(n,n) = \tau_l^2/m_{l,n}$. Matrix $\mathbf{D}$ is referred to here as the diffusion matrix, and is derived in Section V-A1. Instead of this step scheme, we use the one-step Langevin difference equation. To obtain the single step Langevin update equation for $\mathbf{s}_l$, we substitute (17) into (18).

$$\mathbf{s}_l^{k+1} = \mathbf{s}_l^k - \frac{1}{2}\mathbf{D}_l\mathbf{g}(\mathbf{s}_{1:L}^k) + \mathbf{D}_l^{\frac{1}{2}}\mathbf{M}_l^{\frac{1}{2}}\mathbf{v}_l^k \tag{20}$$

This form is also used in [28], [29]. The samples are produced by using this first order equation, and then they are tested in the Metropolis-Hastings scheme.

If we assume the transitions in (20) as a Wiener process and take into account the fact that the velocity vector $\mathbf{v}_l$ is independent of the source vector $\mathbf{s}_l$, [18], then its probability density function can be set as a multivariate Gaussian as $p_{\mathbf{v}_l}(\mathbf{v}_l) = (|\mathbf{M}_l|/2\pi)^{\frac{1}{2}}\exp\left\{-\frac{1}{2}\mathbf{v}_l^T(t)\mathbf{M}_l\mathbf{v}_l(t)\right\}$. We can produce a random sample from this probability such that $\mathbf{v}_l = \mathbf{M}_l^{-\frac{1}{2}}\mathbf{w}_l$ where $\mathbf{w}_l$ is a zero-mean Gaussian vector with identity covariance matrix $\mathcal{N}(\mathbf{w}_l|0,\mathbf{I})$. If we substitute this random sample into (20), we obtain the associated Langevin equation

$$\mathbf{s}_l^{k+1} = \mathbf{s}_l^k - \frac{1}{2}\mathbf{D}_l\mathbf{g}(\mathbf{s}_{1:L}^k) + \mathbf{D}_l^{\frac{1}{2}}\mathbf{w}_l \tag{21}$$

Since the random variables for the image pixel intensities are produced in parallel by using (21), the procedure is faster than the random walk process adopted in [1]. The random walk process produces local random increments independently from the neighbor pixels and the observations. In the Langevin sampler, the samples are generated in an interrelated manner and in terms of the descent of an energy function that reflects the goodness of the model fit. Once the candidate sample image is produced by (21), the accept-reject rule is applied independently to each pixel. In the case of random walk,

TABLE II
METROPOLIS-HASTINGS ALGORITHM FOR A SOURCE IMAGE. u: UNIFORM POSITIVE RANDOM NUMBER IN THE UNIT INTERVAL; $\mathbf{z}$: GENERATED SAMPLE VECTOR TO BE TRIED; $\varphi(z_n, s_{l,n}^k)$: ACCEPTANCE RATIO OF THE GENERATED SAMPLE.

1) $\mathbf{w}_l \sim \mathcal{N}(\mathbf{w}_l|0,\mathbf{I})$
2) $\overline{\mathbf{H}}(\mathbf{s}_l^k) \longleftarrow [\text{diag}\{\mathbf{H}(\mathbf{s}_l)\}_{\mathbf{s}_l \longleftarrow \mathbf{s}_l^k}]^{-1}$
3) $\mathbf{D}_l \longleftarrow 2[\overline{\mathbf{H}}(\mathbf{s}_l^k)]^{-1}$
4) $\mathbf{g}(\mathbf{s}_{1:L}^k) \longleftarrow [\nabla_{\mathbf{s}_l} E(\mathbf{s}_{1:L})]_{\mathbf{s}_{1:L}=\mathbf{s}_{1:L}^k}$
5) produce $\mathbf{z} \longleftarrow \mathbf{s}_l^k - \frac{1}{2}\mathbf{D}_l\mathbf{g}(\mathbf{s}_{1:L}^k) + \mathbf{D}_l^{\frac{1}{2}}\mathbf{w}_l$ from (21).
6) for all pixel $n = 1,\ldots,N$
   a) calculate $\varphi(z_n, s_{l,n}^k)$
   b) if $\varphi(z_n, s_{l,n}^k) \geq 1$ then $s_{l,n}^{k+1} = z_n$
      else produce $u \sim U(0,1)$.
         if $u < \varphi(z_n, s_{l,n}^k)$ then $s_{l,n}^{k+1} = z_n$,
         else $s_{l,n}^{k+1} = s_{l,n}^k$
   c) $n+1 \longleftarrow$ next pixel.

we would produce the candidate sample pixel and apply the accept-reject rule. The sampling of the whole image would be completed by scanning all the pixels in a sequential order as in Gibbs sampling. Since each pixel has to wait the update of the previous pixel, this procedure is very slow. In random walk, candidate pixels can be produced in parallel but, producing a candidate sample for the whole image using random walk is not a reasonable method because hitting the right combination for such a huge amount of data (i.e. $\approx 10^5$) is almost impossible. By Langevin sampler, the likelihood of approximately hitting the right combination at any one step is much higher.

After their production, the samples are tested via Metropolis-Hastings [34] scheme pixel-by-pixel. The acceptance probability of any proposed sample is defined as $\min\{\varphi(s_{l,n}^{k+1}, s_{l,n}^k), 1\}$, where

$$\varphi(s_{l,n}^{k+1}, s_{l,n}^k) \propto e^{-\Delta E(s_{l,n}^{k+1})} \frac{q(s_{l,n}^k|s_{l,n}^{k+1})}{q(s_{l,n}^{k+1}|s_{l,n}^k)} \tag{22}$$

where $\Delta E(s_{l,n}^{k+1}) = E(s_{l,n}^{k+1}, s_{(1:L)-l,n}^k) - E(s_{1:L,n}^k)$ and $E(s_{1:L,n}^k) = W(s_{1:L,n}^k) + U(s_{l,n}^k)$. For any single pixel, $U(s_{l,n})$ can be derived from (13) and (14) as

$$U(s_{l,n}) = \sum_{d=1}^{D} \frac{1+\beta_{l,d}}{2} \log\left[1 + \frac{\phi_d(s_{l,n}, \alpha_{l,d})}{\beta_{l,d}\delta_{l,d}}\right] \tag{23}$$

The proposal density $q(s_{l,n}^{k+1}|s_{l,n}^k)$ is obtained, from (21), as

$$\mathcal{N}\left(s_{l,n}^{k+1}|s_{l,n}^k + \frac{\tau_l^2}{2m_{l,n}} g(s_{1:L,n}^k), \frac{\tau_l^2}{m_{l,n}}\right) \tag{24}$$

One cycle of the Metropolis-Hastings algorithm embedded in the main algorithm, for each source image, is given in Table II.

1) Diffusion Matrix: In this section, we give a method to find an optimum diffusion matrix $\mathbf{D}$. The method must ensure that the produced sample $\mathbf{s}_l^{k+1}$ comes from the joint conditional distribution $p(\mathbf{s}_l, \mathbf{v}_l|\mathbf{y}_{1:K}, \mathbf{s}_{(1:L)-l}, \mathbf{A}, \Theta, \mathbf{M})$ introduced in (12). If we write the Taylor expansion of $E(\mathbf{s}_l^k)$ with the infinitesimal $\Delta\mathbf{s}_l$ and take the expectation of both sides with respect to the joint density $p(\mathbf{s}_l, \mathbf{v}_l|\mathbf{y}_{1:K}, \mathbf{s}_{(1:L)-l}, \mathbf{A}, \Theta, \mathbf{M})$, we obtain the following equation

$$\langle E(\mathbf{s}_l + \Delta\mathbf{s}_l)\rangle = \langle E(\mathbf{s}_l) + \nabla E(\mathbf{s}_l)^T \Delta\mathbf{s}_l + \frac{1}{2}\Delta\mathbf{s}_l^T \mathbf{H}(\mathbf{s}_l)\Delta\mathbf{s}_l\rangle \tag{25}$$

where $\mathbf{H}(\mathbf{s}_l)$ is the Hessian matrix of $E(\mathbf{s}_l)$ with respect to $\mathbf{s}_l$. From this equation, the optimum infinitesimal $\Delta\mathbf{s}_l$ is found as $\Delta\mathbf{s}_l = -[\langle\mathbf{H}(\mathbf{s}_l)\rangle]^{-1}\langle\nabla E(\mathbf{s}_l)\rangle$.

If we also take the expectation of both sides of (20), we obtain

$$\langle\mathbf{s}_l^{k+1}\rangle = \langle\mathbf{s}_l^k\rangle - \frac{1}{2}\mathbf{D}_l g(\mathbf{s}_{1:L}^k) \tag{26}$$

and comparing $\langle\mathbf{s}_l^{k+1}\rangle = \langle\mathbf{s}_l^k\rangle + \Delta\mathbf{s}_l$ with (26), we write $\mathbf{D}_l g(\mathbf{s}_{1:L}^k) = -2[\langle\mathbf{H}(\mathbf{s}_l)\rangle]^{-1}\langle\nabla E(\mathbf{s}_l)\rangle$. Rather than the expectation of the inverse of Hessian matrix, we use its diagonal calculated by the value of $\mathbf{s}_l$ at the discrete time k as

$$\mathbf{D}_l g(\mathbf{s}_{1:L}^k) = -2[\overline{\mathbf{H}}(\mathbf{s}_l^k)]^{-1} g(\mathbf{s}_{1:L}^k) \tag{27}$$

where $\overline{\mathbf{H}}(\mathbf{s}_l^k) = [\text{diag}\{\mathbf{H}(\mathbf{s}_l)\}_{\mathbf{s}_l=\mathbf{s}_l^k}]^{-1}$ and $\text{diag}\{.\}$ operator extract the main diagonal of the Hessian matrix. From (27), we can find the diffusion parameter as [35]:

$$\mathbf{D}_l = 2[\overline{\mathbf{H}}(\mathbf{s}_l^k)]^{-1}. \tag{28}$$

This approximation is justified if $\mathbf{H}(\mathbf{s}_l)$ is strongly diagonally dominant.

B. Parameters of t-distribution

We can write the joint posterior of the parameters $\alpha_{l,d}$, $\beta_{l,d}$ and $\delta_{l,d}$ such that $p(\alpha_{l,d}, \beta_{l,d}, \delta_{l,d}|\mathbf{t}_{l,d}, \Theta_{-\{\alpha_{l,d},\beta_{l,d},\delta_{l,d}\}}) = p(\mathbf{t}_{l,d}|\Theta)p(\beta_{l,d})p(\delta_{l,d})$. Using the likelihood $p(\mathbf{t}_{l,d}|\Theta)$ in (8) and the priors of the parameters, we can find the MAP estimates of the parameters of the t-distribution by EM method. Instead of maximizing the $\log\{p(\mathbf{t}_{l,d}|\Theta)p(\beta_{l,d})p(\delta_{l,d})\}$, we maximize the following function

$$\begin{aligned}\int \log\left\{\frac{p(\mathbf{t}_{l,d}|\Theta)p(\beta_{l,d})p(\delta_{l,d})}{p(\nu_{l,d}|\mathbf{t}_{l,d}^k, \Theta^k)}\right\} p(\nu_{l,d}|\mathbf{t}_{l,d}^k, \Theta^k) d\nu_{l,d} \\ = \langle\log\{p(\mathbf{t}_{l,d}|\Theta)p(\beta_{l,d})p(\delta_{l,d})\}\rangle_{\nu_{l,d}|\mathbf{t}_{l,d}^k,\Theta^k} \\ - \left\langle\log p(\nu_{l,d}|\mathbf{t}_{l,d}^k, \Theta^k)\right\rangle_{\nu_{l,d}|\mathbf{t}_{l,d}^k,\Theta^k}\end{aligned} \tag{29}$$

where $p(\nu_{l,d}|\mathbf{t}_{l,d}^k, \Theta^k)$ is the posterior density of the hidden variable $\nu_{l,d}$ conditioned on parameters estimated in the previous step k and $\langle.\rangle_{\nu_{l,d}|\mathbf{t}_{l,d}^k,\Theta^k}$ represents the expectation with respect to $\nu_{l,d}|\mathbf{t}_{l,d}^k, \Theta^k$. For simplicity, hereafter we use only $\langle.\rangle$ to represent this expectation. The parameter $\nu_{l,d}$ is a hidden (or latent) variable that changes the scale of the Gaussian density $\mathcal{N}(\mathbf{t}_{l,d}|0, \delta_{l,d}^k\mathbf{I}_N/\nu_{l,d})$ and has a Gamma prior $\mathcal{G}(\nu_{l,d}|\beta_{l,d}^k/2, \beta_{l,d}^k/2)$. By exploiting $\nu_{l,d}$, we can define the t-distribution as a scale mixture of Gaussians as in (8). The second term on the righthand side of (29), $-\langle\log p(\nu_{l,d}|\mathbf{t}_{l,d}^k, \Theta^k)\rangle$, corresponds to the entropy of the posterior density of $\nu_{l,d}$, and is independent of the unknowns, and the function

$$Q(\Theta; \Theta^k) = \langle\log\{p(\mathbf{t}_{l,d}|\Theta)p(\beta_{l,d})p(\delta_{l,d})\}\rangle. \tag{30}$$

The aim is to find the maximum of $Q(\Theta;\Theta^k)$ with respect to Θ;

$$\Theta^{k+1} = \arg\max_{\Theta} Q(\Theta;\Theta^k) \tag{31}$$

In the E (expectation) step of the EM algorithm, we must calculate the expectation $\langle . \rangle_{\nu_{l,d}|\mathbf{t}_{l,d}^k,\Theta^k}$. For this purpose, we find the posterior density of $\nu_{l,d}$

$$\begin{aligned} p(\nu_{l,d}|\mathbf{t}_{l,d}^k,\Theta^k) &= p(\mathbf{t}_{l,d}^k|\Theta^k,\nu_{l,d})p(\nu_{l,d}) \\ &= \mathcal{N}(\mathbf{t}_{l,d}^k|0,\delta_{l,d}^k\mathbf{I}_N/\nu_{l,d})\mathcal{G}(\nu_{l,d}|\beta_{l,d}^k/2,\beta_{l,d}^k/2) \\ &= \mathcal{G}\left(\nu_{l,d}|N/2,\phi_d(\mathbf{s}_l^k,\alpha_{l,d}^k)/\delta_{l,d}^k\right)\mathcal{G}\left(\nu_{l,d}|\beta_{l,d}^k/2,\beta_{l,d}^k/2\right) \\ &= \mathcal{G}\left(\nu_{l,d}|\tfrac{N+\beta_{l,d}^k}{2},\tfrac{\beta_{l,d}^k}{2}\left(1+\tfrac{\phi_d(\mathbf{s}_l^k,\alpha_{l,d}^k)}{\beta_{l,d}^k\delta_{l,d}^k}\right)\right). \end{aligned} \tag{32}$$

The expectation of $\nu_{l,d}$ is

$$\langle \nu_{l,d} \rangle = \frac{N+\beta_{l,d}^k}{\beta_{l,d}^k}\left(1+\frac{\phi_d(\mathbf{s}_l^k,\alpha_{l,d}^k)}{\beta_{l,d}^k\delta_{l,d}^k}\right)^{-1} \tag{33}$$

In the M (maximization) step, (30) is maximized with respect to Θ. To maximize this function, we alternate among the variables $\alpha_{l,d}$, $\beta_{l,d}$ and $\delta_{l,d}$. After taking the logarithms and expectations in (30), the cost functions for $\alpha_{l,d}$, $\beta_{l,d}$ and $\delta_{l,d}$ are written as follows

$$Q(\alpha_{l,d};\Theta^k) = -\langle\nu_{l,d}\rangle\frac{\phi_d(\mathbf{s}_l,\alpha_{l,d})}{2\delta_{l,d}} \tag{34}$$

$$Q(\delta_{l,d};\Theta^k) = -\frac{N}{2}\log\delta_{l,d} - \left(\langle\nu_{l,d}\rangle\frac{\phi_d(\mathbf{s}_l,\alpha_{l,d})}{2\delta_{l,d}}\right) \tag{35}$$

$$\begin{aligned} Q(\beta_{l,d};\Theta^k) \quad = \quad & -\log\Gamma(\tfrac{\beta_{l,d}}{2}) + \left(\tfrac{N+\beta_{l,d}}{2}-1\right)\langle\log\nu_{l,d}\rangle \\ & +\tfrac{\beta_{l,d}-1}{2}\log\beta_{l,d} - \tfrac{N+\beta_{l,d}}{2}\log 2 \\ & -\tfrac{\langle\nu_{l,d}\rangle\beta_{l,d}^k}{2}\left(1+\tfrac{\phi_d(\mathbf{s}_l,\alpha_{l,d})}{\beta_{l,d}^k\delta_{l,d}^k}\right) - 0.002\beta_{l,d} \end{aligned} \tag{36}$$

The solutions to (34) and (35) can be easily found as

$$\alpha_{l,d} = \frac{\mathbf{s}_l^T\mathbf{G}_d^T\mathbf{s}_l}{\mathbf{s}_l^T\mathbf{G}_d^T\mathbf{G}_d\mathbf{s}_l} \tag{37}$$

$$\delta_{l,d} = \langle\nu_{l,d}\rangle\frac{\phi_d(\mathbf{s}_l,\alpha_{l,d})}{N} \tag{38}$$

The maximization of (36) does not have a simple solution. It can be solved by setting its first derivative to zero:

$$\begin{aligned} -\psi_1(\tfrac{\beta_{l,d}}{2}) + \log\beta_{l,d} + \langle\log\nu_{l,d}\rangle - \langle\nu_{l,d}\rangle \\ +\tfrac{\beta_{l,d}-1}{\beta_{l,d}} - 0.002 = 0 \end{aligned} \tag{39}$$

where $\psi_1(.)$ is the first derivative of $\log\Gamma(.)$ and it is called digamma function.

C. Parameters of the Mixing Matrix

We assume that the prior of $\mathbf{A}$ is uniform between 0 and ∞. The conditional density of $a_{k,l}$ is expressed as $p(a_{k,l}|\mathbf{y}_{1:K},\Theta^t_{-a_{k,l}}) \propto p(\mathbf{y}_{1:K}|\Theta^t)$. From (2), it can be seen that the conditional density of $a_{k,l}$ becomes Gaussian. The parameter $a_{k,l}$ is estimated in each iteration as

$$a_{k,l} = \frac{1}{\mathbf{s}_l^T\mathbf{s}_l}\mathbf{s}_l^T(\mathbf{y}_k - \sum_{i=1,i\neq l}^{L} a_{k,i}\mathbf{s}_i)u(a_{k,l}) \tag{40}$$

where $u(a_{k,l})$ is the unit step function.

TABLE III
ONE CYCLE OF ADAPTIVE LANGEVIN SAMPLER FOR SOURCE SEPARATION. THE SYMBOL $\longleftarrow$ DENOTES ANALYTICAL UPDATE, THE SYMBOL $\longleftarrow_0$ DENOTES UPDATE BY FINDING THE ZERO ROOT AND THE SYMBOL $\sim$ DENOTES UPDATE BY RANDOM SAMPLING.

Find the initial mixing matrix (i.e. FDCCA [31]).
Find the initial source images using the LS solution.
Initialize the parameters $\alpha_{l,d}^0$, $\beta_{l,d}^0$ and $\delta_{l,d}^0$
for all source images, $l = 1:L$
  for all directions, $d = 1:D$
    $\langle\nu_{l,d}\rangle \longleftarrow \frac{N+\beta_{l,d}^k}{\beta_{l,d}^k}\left(1+\frac{\phi_d(\mathbf{s}_l^k,\alpha_{l,d}^k)}{\beta_{l,d}^k\delta_{l,d}^k}\right)^{-1}$
    $\alpha_{l,d} \longleftarrow \frac{\mathbf{s}_l^T\mathbf{G}_d^T\mathbf{s}_l}{\mathbf{s}_l^T\mathbf{G}_d^T\mathbf{G}_d\mathbf{s}_l}$
    $\delta_{l,d} \longleftarrow \langle\nu_{l,d}\rangle\frac{\phi_d(\mathbf{s}_l,\alpha_{l,d})}{N}$
    $\beta_{l,d} \longleftarrow_0 [-\psi_1(\frac{\beta_{l,d}}{2}) + \log\beta_{l,d} + \langle\log\nu_{l,d}\rangle - \langle\nu_{l,d}\rangle$
    $+\frac{\beta_{l,d}-1}{\beta_{l,d}} - 0.002 = 0]$
  for all pixels, $n = 1:N$
    Using Metropolis-Hastings method in Table II
    $s_{l,n}^{k+1} \sim \left\{p(s_{l,n}|\mathbf{y}_{1:K},\Theta^t_{-s_{l,n}})\right\}$
for all elements of the mixing matrix, $(k,l) = (1,1):(K,L)$
  $a_{k,l} \longleftarrow \frac{1}{\mathbf{s}_l^T\mathbf{s}_l}\mathbf{s}_l^T(\mathbf{y}_k - \sum_{i=1,i\neq l}^{L} a_{k,i}\mathbf{s}_i)u(a_{k,l})$

D. Adaptive Langevin Sampler Algorithm

The proposed Adaptive Langevin Sampler algorithm is given in Table III. The symbol $\longleftarrow$ denotes analytical update, the symbol $\longleftarrow_0$ denotes update by finding the zero root and the symbol $\sim$ denotes the update by random sampling. The sampling of the sources is done by the Metropolis-Hastings scheme given in Table II. To deal practically with uniformly distributed positive variables, we assume that they lie in the range $[0.0001, 1000]$.

1) Initialization: We start the algorithm with the mixing matrix obtained by the FDCCA (Fourier Domain Correlated Component Analysis) [31] method. The initial values of astrophysical maps are obtained by Least Square (LS) solution with the initial mixing matrix. The initial values of $\alpha_{l,d}$ can be calculated directly from image differentials. We initialized the $\beta_{l,d}^0 = 0$ and found the initial value of $\delta_{l,d}$ by equaling the expectation (33) to a constant. In this study, we take the initial value of this posterior expectation 1.5. So the initial value of $\delta_{l,d}^0 = 1.5\phi_d(\mathbf{s}_l^0,\alpha_{l,d}^0)/N$

2) Stopping Criterion: We observe the normalized absolute difference between sequential values of $\mathbf{s}_l$ to decide the convergence of the Markov Chain to an equilibrium. If $|\mathbf{s}_l^k - \mathbf{s}_l^{k-1}|/|\mathbf{s}_l^{k-1}| \leq 10^{-2}$, we assume the chain has converged to the equilibrium for $\mathbf{s}_l$ and denote this point $T_l = k$. Since we have L parallel chains for L sources, the ending point of the burn-in period of the whole Monte Carlo chain is $T_s = \max_l T_l$. We ignore the samples before T_s. We keep the iteration going until T_e that is the ending point of the post burn-in period simulation. In the experiments, we have used 100 iterations after burn-in period, so $T_e = T_s + 100$.

VI. SIMULATION RESULTS

To test our procedure, we assume nine observation channels with center frequencies in the range 30–857 GHz, where the

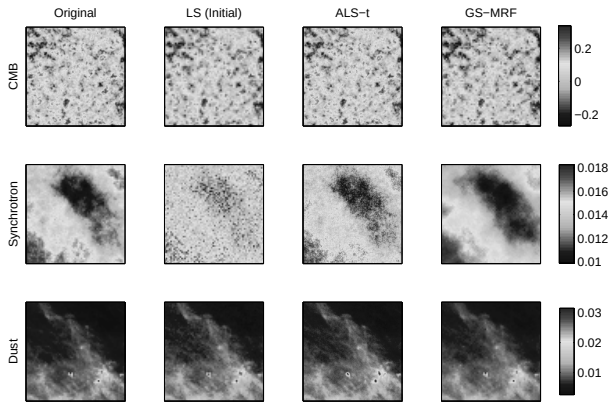


Fig. 3. The separated astrophysical images from noisy observations with the LS, ALS-t and GS-MRF methods. The location of the patch is 0° longitude and 40° latitude, out off galactic plane, and has a size 64×64 pixels.

dominant diffuse radiations are the CMB, the galactic synchrotron radiation and the thermal emission from galactic dust. Except for CMB, these radiations have unknown emission spectra (that is, the coefficients $a_{k,l}$ in (1) are not all known). Both observation models in (1) and (5) are suitable for our algorithm. In the experiments, we assume the model in (1), that is of instantaneous mixtures and isotropic sky, to be valid. We plan to attack the problems of space variability and channel-dependent convolutional effects in the future.

In the sequel, we present astrophysical image separation results on a comparative basis. The proposed method is denoted as ALS-t (Adaptive Langevin Sampler-t-distribution) and is compared to four other methods, namely: 1) GS-MRF, which is the MRF model coupled with Gibbs sampling [1]; 2) LS, which forms our initial estimates on the basis of the values of $a_{k,l}$ obtained by FDCCA [31]; 3) Iterated Conditional Modes (ICM), which maximizes the conditional pdfs sequentially for each variable [11]; 4) ALM-MRF, which is the solution of the MRF model via Langevin and Metropolis-Hastings schemes [22].

The leftmost column of Fig. 3 shows the ground-truth simulated astrophysical source maps. The remaining columns show the source maps separated by LS, ALS-t and GS-MRF, respectively. The sky patch used for this experiment is centered at 0° longitude and 40° latitude in galactic coordinates and has a size of 7.3×7.3 square degrees in the celestial sphere, discretized in a 64×64 pixel map.

The Peak Signal-to-Interference Ratio (PSIR) is used as a numerical performance indicator. The PSIR can be calculated if the ground-truth is known, which is the case in our work since all sky components are simulated. For this patch, the algorithm converges after 155 iterations and uses a total of 255 iterations to reach the solution (see Fig. 4). We compare the results with the ones of LS, ICM, GS-MRF and ALM-MRF.

Table IV lists the PSIR values and the process times. The simulations are run on a Core2 CPU 1.86 GHz PC. The process time of ALS-t is much shorter than that of the GS-MRF. The

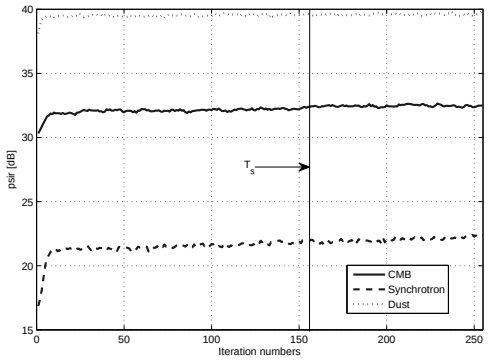


Fig. 4. The PSIRs of the sources in the pixel domain as a function of iteration number. The vertical line signifies the starting point T_s.

TABLE IV
THE PSIR (DB) VALUES OF THE SEPARATED COMPONENTS AND THE PROCESS TIME OF THE ALGORITHMS IN MINUTES.

	CMB	Synchrotron	Dust	time
LS	30.69	15.03	37.37	1.32e-4
ICM	26.27	17.64	35.30	0.31
GS-MRF, [1]	27.81	22.33	38.93	226.72
ALM-MRF, [22]	27.91	20.88	36.41	2.86
ALS-t	33.45	26.21	40.51	1.65

execution time of ALS-t is two orders of magnitude smaller than that of GS-MRF. The PSIR values of ALS-t are also over those of LS, ALM-MRF, GS-MRF and ICM, especially for synchrotron, and furthermore, the smoothing degradation of ICM on the synchrotron component is not observed in the proposed method , Fig. 3.

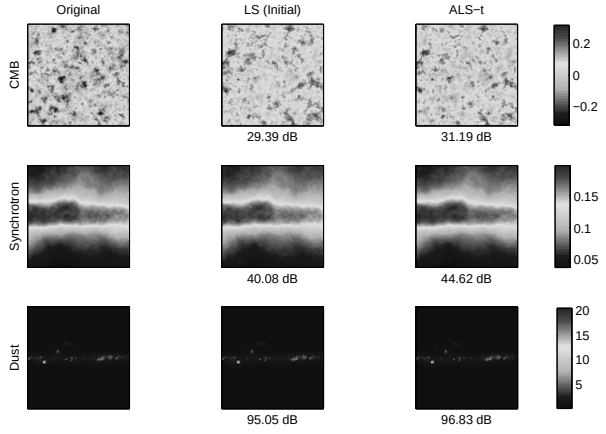


Fig. 5. The separated astrophysical images from noisy observations with the LS and ALS-t methods. The location of the patch is 20° longitude and 0° latitude, galactic plane, and has a size 128×128 pixels.

TABLE V
THE PSIR IMPROVEMENTS (DB) WITH RESPECT TO INITIAL LS SOLUTION.

	CMB	Synchrotron	Dust
$(0^\circ, 40^\circ)$	3.01	10.01	4.08
$(20^\circ, 0^\circ)$	1.80	4.54	1.78

TABLE VI
THE PSIR$_{spec}$ (DB) VALUES OF THE SEPARATED COMPONENTS, IN THE ANNULAR FREQUENCY DOMAIN.

	0°,40°			20°,0°		
	CMB	Synch.	Dust	CMB	Synch.	Dust
LS	30.33	2.65	37.76	31.59	45.69	39.35
ALS-t	35.87	26.23	40.69	34.53	46.98	38.93

We have also run the algorithm for 128×128 pixels patches centered at $(0^\circ, 40^\circ)$ and $(20^\circ, 0^\circ)$. Fig. 5 shows the results for the patch $(20^\circ, 0^\circ)$. In that patch, the relative intensity of CMB is the weakest one. The PSIR values of the estimates of LS and ALS-t are written under the maps and the PSIR improvements are listed for that patch and for the patch $(0^\circ, 40^\circ)$ in Table V. The total time of the ALS-t algorithm for the 128×128 size patches is about 5.31 minutes.

We also use an alternative performance criterion, defined in the spherical harmonic (frequency), ℓ, domain, since the angular power spectrum is relevant to astrophysics. If we decompose a CMB map on spherical harmonics, the complex coefficients, $c_{\ell m}$ ($\ell = 0, 1, 2, \ldots$, $m \in [-\ell, \ell]$), define the angular power spectrum, $C(\ell)$, as the average $C(\ell) = \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} c_{\ell m} c_{\ell m}^*$.

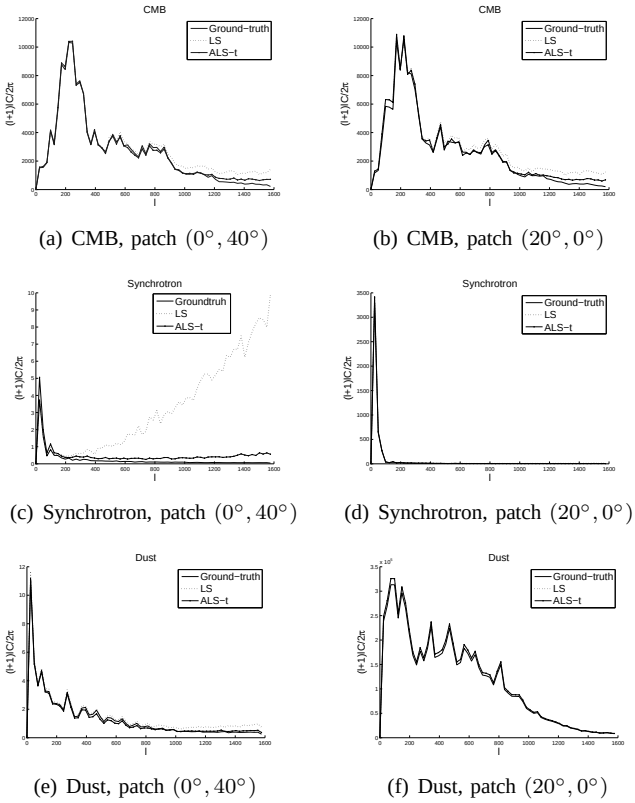
In Fig. 6, we plot the standard power spectrum, $\overline{C}(\ell)$, defined as $\overline{C}(\ell) = (\ell+1)\ell C(\ell)/2\pi$ of the original and the reconstructed sources in the two patches considered. In order to compare different methods, we also introduce the Peak Signal-to-Interference Ratio in the ℓ-domain defined as

$$PSIR_{spec} = 20 \log \left(\frac{\sqrt{\sqrt{N}/2 + 1} \times \max(\overline{C}(\ell))}{||\overline{C}(\ell) - \widehat{\overline{C}}(\ell)||} \right) \quad (41)$$

where $\widehat{\overline{C}}(\ell)$ is the estimated power spectrum.

In the off-galactic patch considered in Fig. 6, the intensity of synchrotron is very low and the LS solution for synchrotron is contaminated too much by noise. The estimated CMB and the dust spectrums by ALS-t follow the ground-truth spectrum better than the LS one, especially in the high frequency regions. For the patch $(20^\circ, 0^\circ)$, synchrotron and dust are estimated adequately by LS, but the LS estimate of CMB is improved by ALS-t. The related PSIR$_{spec}$ values are presented in Table VI.

We have observed that the estimated regression parameter $\alpha_{l,d}$ is quite isotropic for all the maps. For the CMB map in the $(0^\circ, 40^\circ)$ patch, its value is about 0.88 for all d. In the same patch, the values of the parameter $\alpha_{l,d}$ for synchrotron and dust are 0.99. These results show us that the CMB radiation is spatially less correlated than the other radiation sources. We assume the parameter $\beta_{l,d}$ is isotropic and estimate a single value for each direction. The EM estimation of $\beta_{l,d}$ depends too much on its prior and initial value. We have allowed the parameter $\delta_{l,d}$ to be anisotropic, but at the end of the estimation steps we have found that it is almost isotropic for all radiation maps.



(a) CMB, patch $(0^\circ, 40^\circ)$ (b) CMB, patch $(20^\circ, 0^\circ)$

(c) Synchrotron, patch $(0^\circ, 40^\circ)$ (d) Synchrotron, patch $(20^\circ, 0^\circ)$

(e) Dust, patch $(0^\circ, 40^\circ)$ (f) Dust, patch $(20^\circ, 0^\circ)$

Fig. 6. Ground-truth and estimated angular power spectrums for patches $(0^\circ, 40^\circ)$ and $(20, 0)$. The ground-truth spectrum (solid line), spectrum of LS solution (dot line) and spectrum of ALS-t solution (solid line marked with +).

VII. CONCLUSION AND FUTURE WORK

We have developed a Bayesian source separation algorithm for astrophysical images where the MCMC samples are generated through the Langevin stochastic equation. The proposed algorithm provides two orders of magnitude computational economy vis-à-vis the Gibbs sampling approach. In addition, it generates better source separation as compared to all its competitors, i.e., LS, ICM, ALM-MRF and GS-MRF methods measured in terms of PSIR in the pixel domain and PSIR$_{spec}$ in the annular frequency domain. The algorithm can reconstruct the high frequency regions of the power spectrums with higher fidelity. A byproduct of this approach is the capability to estimate the parameters of the t-distribution image priors. Although the proposed ALS-t method takes longer than either ICM or LS methods, its superior performance by far outweighs this disadvantage, and furthermore the algorithm lends itself to parallel processing. To improve the algorithm performance, non-stationary image priors, more efficient discretization time step and diffusion matrix can be investigated in the future. Another point to obtain a beneficial algorithm might be the usage of more than one MH-steps, because the EM algorithm which estimates parameters converges faster than the Monte Carlo sampling scheme.

Our new goal is the application of the proposed algorithm to whole-sky maps. To avoid the difficulties inherent in this problem, we plan to use the "nested numbering" structure provided by the HEALPix [37] package. In this format, we can reach the indexes of the eight neighbors of each pixel on the sphere. To calculate the pixel differences, we will implement a gradient calculation method on the sphere by taking the

non-homogeneous spatial distances between the pixels on the sphere into consideration.

Other issues to be addressed are the channel-dependent blurring effects of the antennas and the non-stationary nature of noise. We have to reformulate the source separation problem without these simplifying assumptions on the observations. Finally, a pixel-based estimation error is being analyzed with the goal of defining a stopping criterion for the algorithm.

ACKNOWLEDGMENT

The simulated source maps (see [36]) were provided by the Planck technical working group on diffuse component separation (WG2.1). Some of the results in this paper have been derived using the HEALPix package [37].

REFERENCES

[1] K. Kayabol, E. E. Kuruoglu and B. Sankur, "Bayesian separation of images modelled with MRFs using MCMC," *IEEE Trans. Image Process.*, vol.18, no.5, pp. 982–994, May 2009.

[2] K. Kayabol, and E. E. Kuruoglu, and B. Sankur, "Image Source Separation using Color Channel Dependencies," in *ICA 2009, LNCS*, vol. 5441, pp. 499–506, Paraty, Brasil, Springer-Verlag, 2009.

[3] J.-F. Cardoso, "The three easy routes to independent component analysis; Contrasts and Geometry," in *Int. Conf. on Indepen. Comp. Anal. ICA'01*, San Diego, Dec. 2001.

[4] Student (W. S. Gosset), "The probable error of a mean," *Biometrika*, vol.6, no.1, pp. 1–25, 1908.

[5] D. Higdon, *Spatial Applications of Markov Chain Monte Carlo for Bayesian Inference*. PhD Thesis, University of Washingthon, 1994.

[6] I. Prudyus, S. Voloshynovskiy and A. Synyavskyy, "Wavelet-based MAP image denoising using provably better class of stochastic i.i.d. images models," in *Int. Conf. on Telecomm. in Modern Satell., Cable and Broadcas. TELSIKS'01*, pp. 583–586, Sep. 2001.

[7] G. Chantas, N. Galatsanos, A. Likas and M. Saunders, "Variational Bayesian image restoration based on a product of t-distributions image priors," *IEEE Trans. Image Process.*, vol.17, no.10, pp. 1795–1805, Oct. 2008.

[8] D. Tzikas, A. Likas and N. Galatsanos, "Variational Bayesian sparse kernel-based blind image deconvolution with Student's-t priors," *IEEE Trans. Image Process.*, vol.18, no.4, pp. 753–764, Apr. 2009.

[9] C. Fevotte and S. J. Godsill "A Bayesian approach for blind separation of sparse sources," *IEEE Trans. Audio, Speech Language Process.*, vol. 14, no. 6, pp. 2174–2188, Nov. 2006.

[10] T. Hebert and R. Leahy, "A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors," *IEEE Trans. Medical Imaging*, vol.8, no.2, pp. 194–202, June 1989.

[11] D. B. Rowe, "A Bayesian approach to blind source separation," *Journal of Interdisciplinary Mathematics*, vol.5, no.1, 2002.

[12] K. H. Knuth, "A Bayesian approach to source separation, " in *Int. Conf. on Indepen. Comp. Anal. ICA'99*, pp. 283–288, Jan. 1999.

[13] A. Mohammad-Djafari, "A Bayesian approach to source separation," in *Int. Workshop on Maximum Entropy and Bayesian Methods, MaxEnt'99*, July, 1999.

[14] A. Tonazzini, L. Bedini, and E. Salerno "A Markov model for blind image separation by a mean-field EM algorithm," *IEEE Trans. Image Process.*, vol. 15, pp. 473–482, Feb. 2006.

[15] E. E. Kuruoglu, A. Tonazzini, and L. Bianchi, "Source separation in noisy astrophysical images modelled by Markov random fields," in *Int. Conf. on Image Proc. ICIP'04*, pp. 24–27, Oct., 2004.

[16] P. Langevin, "Sur la theorie du mouvement brownien," (On the theory of Brownian motion), *C.R. Acad. Sci.*, (Paris), vol. 146, pp. 530-533, 1908.

[17] R.J. Rossky, J.D. Doll, and H.L. Friedman, "Brownian dynamics as a smart Monte Carlo simulation," *J. Chem. Phys.*, vol. 69, pp. 4628–4633, 1978.

[18] R.M. Neal, "Probabilistic inference using Markov chain Monte Carlo methods," Tech. Rep. CRG-TR-93-1, Dept. Comp. Scien., University of Toronto", Sep. 1993.

[19] D.M. Higdon, J.E. Bowsher, V.E. Johnson, T.G. Turkington, D.R. Gilland, and R. J. Jaszczak", "Fully Bayesian estimation of Gibbs hyperparameters for emission computed tomography data," *IEEE Trans. Medical Imaging*, vol. 16, no. 5, pp. 516–516, Oct. 1997.

[20] X. Descombes, R.D. Morris, J. Zerubia, and M. Berthod, "Estimation of Markov random field prior parameters using Markov chain Monte Carlo maximum likelihood," *IEEE Trans. Image Process.*", vol. 8, no. 7, pp. 954–963, July 1999.

[21] R. Molina, A. K. Katsaggelos and J. Mateos, "Bayesian and regularization methods for hyperparameter estimation in image restoration," *IEEE Trans. Image Process.*", vol. 8, no. 2, pp. 231–246, Feb. 1999.

[22] K. Kayabol, E. E. Kuruoglu, B. Sankur, E. Salerno and L. Bedini, "Fast MCMC separation for MRF Modelled astrophysical components," in *Int. Conf. on Image Proc. ICIP'09*, pp. 2769–2772, Nov. 2009.

[23] Planck Science Team, "PLANCK: The scientific programme," *European Space Agency (ESA)*, 2005. [Online]. Available: http://www.esa.int/SPECIALS/Planck/index.html

[24] W. Hu and S. Dodelson, Cosmic Microwave Background Anisotropies, *Annual Review of Astronomy and Astrophysics*, vol. 40, pp. 171-216, 2002.

[25] C. Baccigalupi, L. Bedini, C. Burigana, G. De Zotti, A. Farusi, D. Maino, M. Maris, F. Perrotta, E. Salerno, L. Toffolatti, A. Tonazzini "Neural networks and the separation of cosmic microwave background and astrophysical signals in sky maps," *Mon. Not. Royal Astronom. Soc.*, vol.318, pp. 769-780, 2000.

[26] C. A. Bonaldi, L. Bedini, E. Salerno, C. Baccigalupi, and G. De Zotti, "Estimating the spectral indices of correlated astrophysical foregrounds by a second-order statistical approach," *Mon. Not. Royal Astronom. Soc.*, vol.373, 271-279, 2006.

[27] G. E. Hinton,"Products of experts," in *Int. Conf. on Artificial Neural Net. ICANN'99*, vol. 1, pp. 1–6, 1999.

[28] U. Grenander and M. I. Miller, "Representations of knowledge in complex systems (with discussion), " *J. R. Statist. Soc. B*", vol. 56, pp. 549–603, 1994.

[29] P. Dostert, Y. Efendiev, T. Y. Hou and W. Luo, "Coarse-gradient Langevin algorithms for dynamic data integration and uncertainty quantification, " *J. Comput. Phys.*", vol. 217, pp. 123–142, 2006.

[30] C. Liu and D. B. Rubin, "ML estimation of the t distribution using EM and its extensions, ECM and ECME," *Statistica Sinica*", vol. 5, pp. 19–39, 1995.

[31] L. Bedini, and E. Salerno, "Extracting astrophysical source from channel-dependent convolutional mixtures by correlated component analysis in the frequency domain," in *Lecture Notes in Artificial Intelligence*, vol. 4694, pp. 9–16, Springer-Verlag, 2007.

[32] National Aeronautics and Space Administration, "Cosmic Background Explorer," *NASA*. [Online]. Available: http://lambda.gsfc.nasa.gov/product/cobe/

[33] National Aeronautics and Space Administration, "Wilkinson Microwave Anisotropy Probe," *NASA*. [Online]. Available: http://map.gsfc.nasa.gov/

[34] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*", vol. 57, no. 1, pp. 97–109, Apr. 1970.

[35] S. Becker and Y. Le Cun, "Improving the convergence of back-propagation learning with second-order methods," in *Proc. of the 1988 Connectionist Models Summer School*, pp. 29–37, 1989.

[36] L. Bedini, D. Herranz, E. Salerno, C. Baccigalupi, E. Kuruoglu, A. Tonazzini, "Separation of correlated astrophysical sources using multiple-lag data covariance matrices", Eurasip J. on Appl. Sig. Proc., vol. 2005, no. 15, pp. 2400-2412, Aug. 2005.

[37] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, M. Bartelmann, "HEALPix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere", *The Astrophysical Journal*, vol 622, Issue 2, pp. 759-771, 2005.

Koray Kayabol (S'03, M'09) was born in Sakarya, Turkey in 1977. He received the B.Sc., M.Sc. and Ph.D. degrees in electrical&electronics engineering from Istanbul University, Istanbul, Turkey in 1997, 2002 and 2008, respectively.

He was a research assistant in Electrical & Electronics Eng. Dept. between 2001 and 2008. Since 2008, he has been with the ISTI-CNR, Pisa, Italy as a postdoctoral researcher. His research interests include Bayesian image processing and statistical image models.

PLACE PHOTO HERE

Ercan Kuruoğlu (SM'06, M'98) was born in 1969 in Ankara, Turkey. He received his PhD degree from the University of Cambridge in 1998.

He joined the Xerox Research Center, Europe in Cambridge in 1998. He was an ERCIM Fellow in 2000 in INRIA-Sophia Antipolis, France. In January 2002, he joined ISTI-CNR, Pisa. He was a visiting professor in Georgia Tech-Shanghai in Autumn 2007. He is currently a Senior Researcher at ISTI-CNR. His research interests are in the areas of statistical signal and image processing and information and coding theory with applications in astrophysics, geophysics, bioinformatics and telecommunications.

He was an Associate Editor for IEEE Transactions on Signal Processing in 2002-2006 and for IEEE Transactions on Image Processing in 2005-2009 and is in the editorial boards of Digital Signal Processing: a Review Journal and EURASIP Journal on Advances in Signal Processing. He acted as the technical chair for EUSIPCO 2006. He is a member of the IEEE Technical Committee on Signal Processing Theory and Methods.

PLACE PHOTO HERE

José Luis Sanz received the Ph.D. degree in theoretical physics from Universidad Autonoma de Madrid, Spain, in 1976. He was a M.E.C. Postdoctoral Fellow at the Queen Mary College, London, U.K., during 1978. He is currently at the Instituto de Física de Cantabria, Santander, Spain, as UC Professor on Astronomy since 1987.

His research interests are in the areas of Cosmic Microwave Background astronomy (anisotropies, non-Gaussianity), extragalactic point sources and clusters of galaxies (blind/non-blind detection, estimation, statistics) as well as the development of techniques in signal processing (wavelet design, linear/non-linear filters, time-frequency, sparse representations) and application of such tools to astronomical data.

PLACE PHOTO HERE

Bülent Sankur has received his B.S. degree in Electrical Engineering at Robert College, Istanbul, and completed his graduate studies at Rensselaer Polytechnic Institute, New York, USA. His research interests are in the areas of Digital Signal Processing, Image and Video Compression, Biometry, Cognition and Multimedia Systems. He has established a Signal and Image Processing laboratory and has been publishing 150 journal and conference articles in these areas.

Since then he has been at Boğaziçi (Bosporus) University in the Department of Electric and Electronic Engineering. He has held visiting positions at University of Ottawa, Technical University of Delft, and Ecole Nationale Suprieure des Tlcommications, Paris. He also served as a consultant in several private and government institutions.

Prof. Sankur is serving in the editorial boards of three journals on signal processing. He was the chairman of ICT'96: International Conference on Telecommunications and EUSIPCO'05: The European Conference on Signal Processing as well as technical chairman of ICASSP'00.

PLACE PHOTO HERE

Emanuele Salerno graduated in electronic engineering from the University of Pisa, Italy, in 1985. In 1987, he joined the Italian National Research Council as a full-time researcher. At present, he is a senior researcher at the Intitute of Information Science and Technologies in Pisa, Signals and Images Laboratory.

He has been working in applied inverse problems, image reconstruction and restoration, microwave nondestructive evaluation, and blind signal separation, and held various responsibilities in research programs in nondestructive testing, robotics, numerical models for image reconstruction and computer vision, and neural network techniques in astrophysical imagery. He has been supervising various theses in Computer Science, Electronic and Communications Engineering, and Physics, and, since 1996, has been teaching courses at the University of Pisa.

Dr. Salerno is a member of the Italian society for information and communication technology (AICT-AEIT).

PLACE PHOTO HERE

Diego Herranz received the B.S. degree in 1995 and the M.S. degree in physics from the Universidad Complutense de Madrid, Madrid, Spain, in 1995 and the Ph.D. degree in astrophysics from Universidad de Cantabria, Santander, Spain, in 2002. He was a CMBNET Postdoctoral Fellow at the Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" (CNR), Pisa, Italy, from 2002 to 2004. He is currently at the Instituto de Física de Cantabria, Santander, Spain, as UC Teaching Assistant.

His research interests are in the areas of Cosmic Microwave Background astronomy and extragalactic point source statistics as well as the application of statistical signal processing to astronomical data, including blind source separation, linear and nonlinear data filtering, and statistical modeling of heavy-tailed processes.

This figure "herra.jpg" is available in "jpg" format from:

http://arxiv.org/ps/1101.1396v1

This figure "kayab.jpg" is available in "jpg" format from:

http://arxiv.org/ps/1101.1396v1

This figure "kuruo.jpg" is available in "jpg" format from:

http://arxiv.org/ps/1101.1396v1

This figure "saler.jpg" is available in "jpg" format from:

http://arxiv.org/ps/1101.1396v1

This figure "sanku.jpg" is available in "jpg" format from:

http://arxiv.org/ps/1101.1396v1

This figure "sanz.jpg" is available in "jpg" format from:

http://arxiv.org/ps/1101.1396v1